

# Grand Challenge Award: Interactive Visual Analytics

## Palantir: The Future of Analysis

Jason Payne     Jake Solomon     Ravi Sankar     Bob McGrew

### ABSTRACT

Palantir is a world-class analytic platform used worldwide by governmental and financial analysts. This paper provides an introduction to the platform contextualized by its application to the 2008 IEEE VAST contest. In this challenge, we explored a notional dataset about a fabricated religious movement, Catalano's Paraiso Manifesto Movement.

**KEYWORDS:** Palantir, VAST 2008, visual analytics, data visualization, collaboration

**INDEX TERMS:** H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

### 1 INTRODUCTION

Today, data is a commodity. Raw data is growing more abundant and inexpensive every day. Unfortunately, these new datasets are frequently extremely large and contain only a kernel of significant information. Some have tried to throw automated processing at the search for relevance, but computers pale in comparison to humans at this task.

Our goal is to allow analysts to uncover relevance by asking rich, high-level questions of their data without having to know languages such as SQL or think like a computer. From the outset, Palantir was designed with the input of operators and analysts in the field. Their invaluable feedback has enabled the creation of a product that allows users to do more work, in less time, while also providing a higher level of analytic insight.

### 2 THE PALANTIR PLATFORM

Palantir is the market-leading analytical platform used in both the Financial and Government communities. By combining a powerful backend with an intuitive frontend, Palantir enables customers to execute complex analytic and investigative missions with greater speed and fewer resources.

Our best-of-breed application is developed by a team of dedicated Silicon Valley engineers. Because of the enterprise nature of complex analytical problems, Palantir is server/client software. Palantir integrates seamlessly with legacy systems, features numerous APIs, and uses standard XML formats for both import and export.

The Palantir is based on three platforms:

- Data Platform
- Application Platform
- Compute Platform

#### 2.1 DATA PLATFORM

The Data Platform's power derives from its ability to import raw data (structured and unstructured), as well as ontological or relational semantics, via application programming interfaces (API's).

Palantir provides a pluggable Extraction & Transformation framework so that data from any source can be imported into a Palantir instance or federated against

Integrating various data resources into a given user's analytical environment creates the most sophisticated Data Platform for enterprise-level relational analysis.

Data can be moved into and out of a Palantir instance using a W3C compliant xml format that is a truly lossless representation of the Palantir data model. This standardized format provides a valuable way to allow Palantir data to interact with any other open file format.

#### 2.2 APPLICATION PLATFORM

The Palantir frontend is natively enabled with various applications (e.g. the Graph) and applications helpers (e.g. the Histogram).

In addition, Palantir provides frontend extensibility points that allow third parties to build their own applications or embed current tools directly in the Palantir Platform. Both applications and their helpers easily integrate with the Application Platform using Palantir's extensive Application, Helper, EEWS, and AuthWS APIs.

#### 2.3 COMPUTE PLATFORM

Computational tasks like sorting, searching, and database indexing present particular barriers in generating (1) fast and (2) resource-minimal results. This becomes all the more pertinent when compute clusters can range from hundreds or thousands of processors.

The Palantir Compute Platform intelligently distributes and parallelizes tasks to nodes in an Enterprise's computing network.

This MapReduce-Based, distributed Platform lets Palantir take advantage of parallel processing, where the Platform can run on thousands of machine clusters with high scalability

### 3 VAST AND PALANTIR: THE ANALYTIC PROCESS

With Palantir, statistical, relational, temporal, geospatial and Social Network Analyses blend seamlessly within the same platform.

The first step in analyzing the data from the VAST contest was to build an ontology to account for the data from the 4 challenges. Palantir is agnostic to the underlying data though a technology called the Dynamic Ontology that allows the user to define their objects, properties, events, and links/relationships. We created events like Landings and Wiki Edits, entities like Persons and Cell Towers, and properties like Usernames and Vessel Type. We then imported the four data sources provided.

Once the data has been imported, the analyst brings some initial data to the Graph, enabling both relational and social network analyses. Analysis paths from VAST included known landings of migrant boats, first and second order cell phone network of a suspect, or all Wiki edits—whatever information seems like a logical launching point for the investigation at hand.

---

100 Hamilton Ave Suite 300 Palo Alto CA 94301  
inquiries@palantirtech.com

IEEE Symposium on Visual Analytics Science and Technology  
October 19 - 24, Columbus, Ohio, USA  
978-1-4244-2935-6/08/\$25.00 ©2008 IEEE

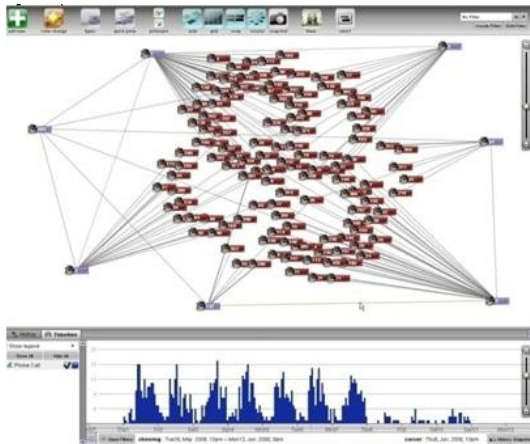


Figure 1: A silenced communication network

With relevant information on graph, the next step is to use the Histogram for statistical analysis. For example, we can quickly determine which users are making the most edits on Wikipedia, which cell phone IDs are involved in the most calls, or how many boats are landing during a selected period of time.



Figure 2: Vessel Type of Mexican Landings

The Timeline enables temporal analysis by graphically representing a histogram of events over time. It is interactive, allows us to isolate interesting patterns or to find spikes/low-points in activity.

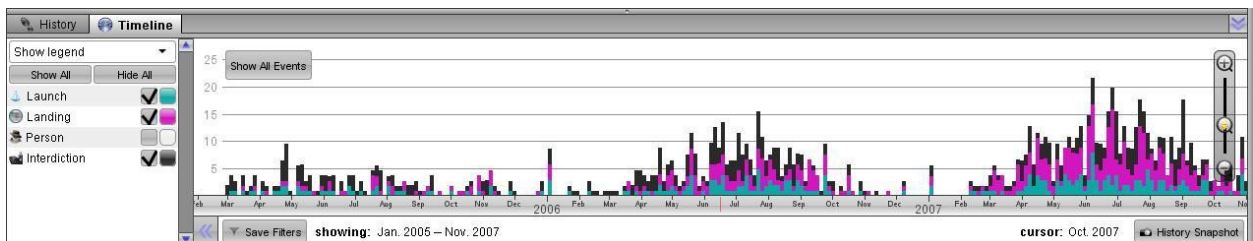


Figure 3: Launches, Landings, and Interdictions in the Palantir Timeline

Lastly, geospatial analysis is enabled by the Palantir data model. Every property in Palantir has geotime metadata – time range and geocords. Because of the open architecture of Palantir, Lat/Long, MGRS, and UTM can be interchanged, and geospatial data can be exported as Shape or KMZ to any leading geospatial product.



Figure 4: Landings on the Southern tip of Florida

#### 4 CONCLUSION

Palantir is a powerful enterprise investigative platform enabling analysts to uncover relationships and patterns hidden within disparate data sources. Our experience with the VAST 2008 competition has reinforced our belief in the need for analytic *platforms*, not *products*. Openness and flexibility are key to being able to cope with the demands of a broad range of analytic tasks. In addition to our existing open APIs and open XML formats, our pluggable architecture enables users to design custom analysis tools and helpers that will be seamlessly integrated into the analysts' workspace, enabling advanced visualization solutions for increasingly complex analytical challenges.